Authors reply to

Journal: Climate of the Past
Title: Statistical downscaling of a climate simulation of the last glacial cycle: temperature and precipitation over Northern Europe
Authors: N. Korhonen, A. Venäläinen, H. Seppä, H. Järvinen
MS No.: Clim. Past Discuss., 9, 3371–3398, 2013
MS Type: Research Article
Referee #2

We would like to thank the Referee for the constructive and good comments. Below are our replies to the comments

General comments

The basic principle appears right - selecting a model by trying to get the best fit for a given number of degrees of freedom. However, the choice will be among the tested predictors, and while it is interesting, I am surprised with what came up as equations 3, 4, and 5.

Comment: All have a sort of "map", the function of latitude and longitude sj(x, y). For the method to be valid, any contribution of these maps to the final result needs to be fully independent of time: it needs to be limited to the correction of a sort of "local bias" in the interpolated EMIC result. The potential to achieve this probably depends on the "diversity" in the calibration data: in the extreme case which would only use one calibration time-slice, such a map could, depending on the available degrees of freedom, represent most of the data. We are evidently not in this case, as there are 2 glacial time-slices + the recent past: the risk of exaggerating the contribution of this time-independent term (which might not be time-independent in the real world) is limited. However, this suggests that the model needs to be carefully validated, because there is a potential to obtain a very good fit for the calibration data, but much less good results when predicting other time periods - even in the "calibration range" of each predictor. A related difficulty is that the model includes quite complex functions, in particular s7(x, y, TCLI). This particular contribution (s7) suggests that some locations are more "sensitive" to the EMIC's output than others (could an interpretation be provided?). This is perhaps true, but again, there is a potential for this contribution to appear independent of time for the calibration cases, while it is not clear that it is more generally true (in the real world).

Reply: We tested about 50 combinations of predictors and found the presented best.
All the predictors were physically reasonable.
Due to the northern location we tested the effect of including the term distance to the ice sheet.

To widen the validation of the models we tested the monthly GAMs in predicting other months' temperature or precipitation. The annual GAMs (calibrated only with LGM and recent past climate data) were tested to predict the 44 kyr BP annual mean temperatures and precipitation and the results were compared to simulations with RCA3. These results are given in supplementary material.

Comment: The formulation of the GAM used for precipitation is relatively surprising. It appears that the model for log (P) is the addition of splines including the precipitation from the EMIC, PCLI (would log(PCLI) provide the same results?).

Reply: Yes, in our tests, the log(P_{CLI}) did provide the same results.

Comment: Thus, Figure 2(a) suggest that for high (monthly) precipitation, the expected regional rainfall includes a term increasing exponentially with the precipitation from the EMIC, while for medium precipitation amounts; there is relatively little influence of the precise EMIC simulated value. It seems possible but not evident that this will remain valid under different climate conditions. It does not seem

easy to provide a clear interpretation for the results in figure 2, especially panels a) and c), therefore this also suggests that the validation should be very careful.
However, I do not have the impression that the validation is sufficiently careful in the current version of the manuscript: was there an attempt at evaluating the GAM by comparing its results to the observations that were not included in the calibration data? I apologize if I missed something - I did not find such a validation in the manuscript. It could probably be based on 1) calibrating the model on the "present" and LGM data, then comparing it to the 44 kyr data, or 2) calibrating the model on some month(s) and validating it on other(s). Achieving good results with (1) would be impressive. Could such a validation be added? (if something is already done, then please clarify what is the data used for calibration and what is the data used for validation).

Reply: Here we added some tests which compared monthly precipitation GAMs and summer monthly temperature GAMs predictions to other months. The annual GAMs were tested to predict the 44 kyr BP annual mean temperature and precipitation and the results were compared to simulations with RCA3. The results are as supplementary material.

Comment: The calibration method would probably benefit from a clarification - is the calibration performed independently for each month, using monthly values for each grid point and all the 3 time slices?

Reply: Yes, for the monthly GAMs the calibration was performed independently for each month using monthly values for each grid point at all the three time steps. This has now been explained in the text.

Comment: In summary, I do not have the impression that the results from tables 1 and 2, as well as figures 3 and 5, are currently sufficiently convincing regarding the validation of the statistical model: they are interesting as they show that an acceptable fit is achieved for the calibration period, but it would be useful to find a way to confirm that the choice of predictors is appropriate and that the GAMs do provide good predictions for other periods.

Reply: See added calibration tests

Comment: In the current version, the most important result in this regard could be figure 6. However, I do not understand why the main difference seems to be a systematic error (bias) over all the period. Are the GAMs performing better than a simple correction of the present-day bias in the interpolated result from CLIMBER? Again, I apologize if I am missing something - I would feel it strange that the comparison can be completely changed by a simple bias correction and that it was not done.

Reply: Based on Figure 6, it seems at first that the GAM is simply a bias correction of CLIMBER simulation towards the paleo-reconstructions. The bias of the reconstructions is however unknown and thus we cannot definitely say that the GAM is an improvement over CLIMBER only because it is closer to the reconstructions (we of course hope it is, but we cannot say). Bearing in mind that the reconstructions are independent of the GAM and CLIMBER, it is nevertheless satisfying to see that the GAM is in its proximity. If we believe in the reconstructions, then we can conclude that the additional information provided to the GAM, on top of the CLIMBER simulation, is adequate and very general since the locations of the reconstructions are completely random for the GAM, and we have all reasons to further believe that the GAM would fit reconstructions is other locations, as well. Unfortunately, there are extremely few sites with such data available.

Comment: The changes shown on figure 6 also appear relatively small; hence the potential to compare the observations to the statistical model is limited: could more locations and/or a longer period be provided?

Reply In the planned next work we aim to make a more comprehensive comparison with the paleoclimatic observations. This work has mainly concentrated in the introduction of the developed method.

Specific comments

Comment: Page 3376 - line 12: I do not understand the sentence referring to a "stepwise screening of the data".

Reply: Confusing sentence can be deleted

Comment: Equation (1): Remark: +_ is probably not required (as this expression provides the expected value, including an error term (residuals) does not seems appropriate, please check)

Reply: Yes, $+\varepsilon$ is not required in the equation (Wood 2006, p. 121)

Comment: page 3377 - line 6 - 9: How can predictors be extracted from the RCA and CCSM models ? I though that those models were used to calibrate the GAMs?

Reply: They are extracted for the calibration only. For predicting with the GAM model, the predictors are extracted from CLIMBER-2 and SICOPOLIS output.

Comment: Equation (2): Please add a reference, including for the "cost weight" = 1.4

Reply: reference Kim and Gu, 2004.

Kim, Y.J. and C. Gu (2004). Smoothing spline Gaussian regression: more scalable computation via efficient approximation. Journal of the Royal Statistical Society, Series B 66, 337-356.

Comment: Equation (6), (7) and (8): why are there different notations for expected values – what is the difference between the overline in equation 6 and the brackets in equation 8, and what are the differences between the definitions provided in equations 6 and 8 ?

Reply: Yes, the equation (8) is not needed as we have (6). the (8) could be left out and the (7) could be written as:

$$Cor = \frac{\sum_{i=1}^{n}(Y_m - \overline{Y_m})(Y_0 - \overline{Y_0})}{\sqrt{\sum_{i=1}^{n}(Y_m - \overline{Y_m})^2 \sum_{i=1}^{n}(Y_0 - \overline{Y_0})^2}}$$

Comment: page 3383 - line 3: I would expect the wording "skill scores" to apply to model "predictions", that is, not for the calibration cases. If this is about predictions, please clarify; otherwise I think that the wording should be changed.

Reply: Yes, the wording could be just "skills

Comment: page 3383 - line 14: The temperature change with altitude seems rather small, at ~2°C/km. Could you comment?

Reply: The CLIMBER model does have a coarse elevation grid too. Therefore the GAM model does not need to correct all the temperature differences, here for Northern Europe the GAM is a lot calibrated with the large ice sheet which already is noticed by CLIMBER to some extent, i.e., the GAM does correct only a part of the temperature bias by altitude.

Comment: page 3384 - line 11: What is meant by "general error" ? Is it systematic or random?

Reply: More detailed text is "bootstrapping-based root mean squared errors of prediction for the reconstructed values are sample-specific and vary from 1.0 to 1.5 °C"

Comment: page 3385 - line 27: Remark: be careful regarding simulations for the future, as it would require a different calibration and validation, including simulations with much more GHGs.

Reply: We can leave out this sentence regarding to simulations for the future.

Comment: References: please check the status of the papers mentioned as "submitted", in particular "Martin et al., 2013a, 2013b, 2013c". If these are not accepted, they should not be used as references.

Reply: The papers Martin et al., 2013a, 2013b, 2013c should be deleted from the Reference list and they shall not be sited anymore in the manuscript.